User Manual

TextSoap® 5
Total Text Transformation

From unmarked software

TextSoap® 5

Acknowledgements

Portions of this software product may utilize the following copyrighted material, the use of which is hereby acknowledged:

Isao Sonobe – (OgreKit)
Copyright © 2006 Isao Sonobe <sonobe@gauge.scphys.kyoto-u.ac.jp>
All rights reserved.

K.Kosako – (OniGuruma)
Copyright © 2006 K.Kosako <kosako@sofnec.co.jp>
All rights reserved.

Matt Gemmell – (RoundedBox)
Copyright © 2005, Matt Gemmell. All rights reserved.
http://mattgemmell.com/

Feel free to use this code in your software, either as-is or in a modified form. Either way, please include a credit in your software's "About" box or similar.

Matt Gemmell – (RoundedBox)
Copyright © 2005, Matt Gemmell. All rights reserved.
http://mattgemmell.com/

Troy N. Stephens – (NSString - getFSRef:)
Copyright © 2001-2006, Troy N. Stephens.

A complete list of acknowledgements can be found under "About TextSoap" within the application.

Table of Contents

This page intentionally left blank.

Getting Started

TextSoap is a powerful text transformation tool. With it, you can remove extraneous characters, rewrap text, or perform one of more than 80 different actions (not including your own) with a single click. You can apply TextSoap to any type of text document that might normally require tedious finding and replacing. TextSoap is often applied on E-mail text but is not restricted to them.
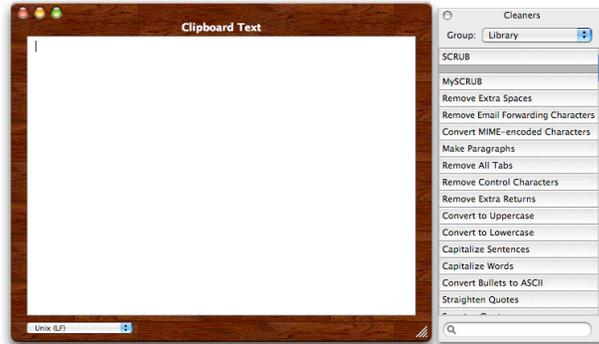
A common working environment of TextSoap consists of the Clipboard Text and the Cleaner list.

TextSoap also enables you to integrate any text document directly into the application itself or to use TextSoap as an integrated plug-in of various supported applications.

There are three basic steps to using TextSoap:

1. Select the text to clean
2. Apply Cleaner
3. Replace text

Some examples of applying a few common cleaners will be looked at in the next section.

Text Cleaners 101

What is a text cleaner and what does it do? The term "cleaner" is used by TextSoap to represent its text transformation steps. Each cleaner includes one or more transformation steps that are applied to the text.

"Convert to Uppercase" is one of the more basic cleaners. It converts the given text to all uppercase.

Here is how you do it: select the text you want cleaned (I used an example out of the manual),

Highlight the text in the Clipboard. Now click on the desired cleaner, for example "Convert to Uppercase".

Now all characters are converted to uppercase.

We have all uppercase letters but might still have those annoying forwarding marks (the greater than symbol '>'). This time, we can click on "SCRUB". "SCRUB" addresses 90% of your text cleaning needs.
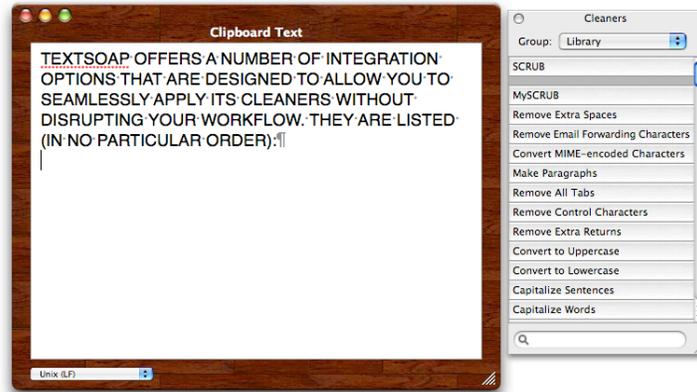
"SCRUB" is an example of a more advanced cleaner. It applies several different cleaners in a specific order to remove the most common garbage characters and formatting problems encountered. Scrub applies the cleaners in the proper order to get the best possible results.

TextSoap includes more than 80 built-in cleaners plus the ability to create your own. You may not need to use all the cleaners that are provided, depending on the particular task you are engaged in, but they are available if and when you do need them.

TextSoap also offers the ability to group cleaners. There are several pre-defined groupings, including MyList.

TextSoap Standard users can change which cleaners are listed in the MyList group. TextSoap Deluxe users can create a virtually unlimited number of additional groups.
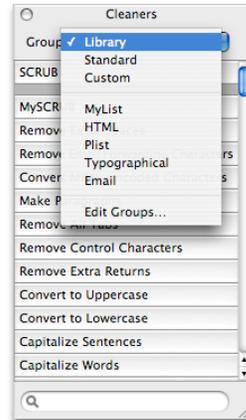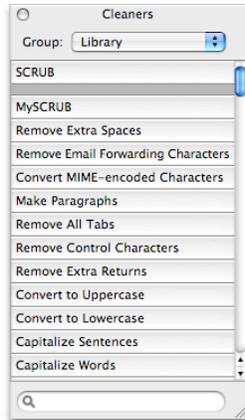
TextSoap Overview

TextSoap offers a number of integration options that are designed to allow you to seamlessly apply its cleaners without disrupting your workflow. They are listed (in no particular order):

Option 1: Application

The following steps describe one way of using TextSoap, these steps will work on almost every type of text documentation.

1. Copy text to clipboard from your application
2. Launch TextSoap application
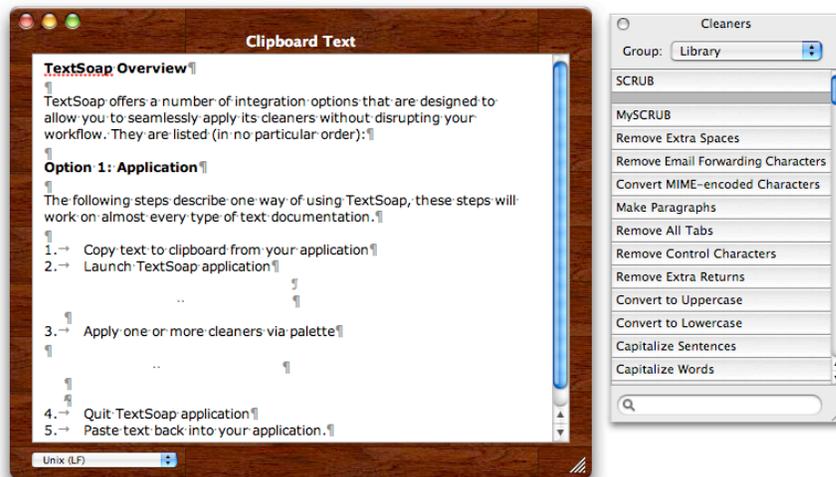
3. Apply one or more cleaners via palette
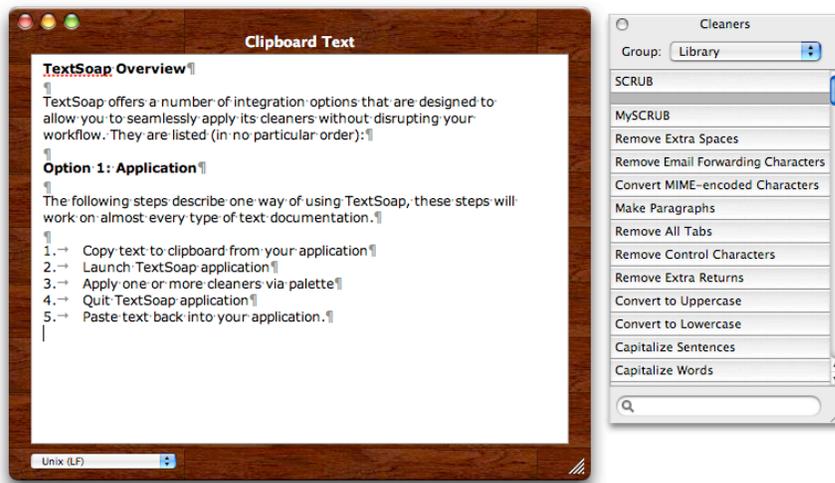
4. Quit TextSoap application
5. Paste text back into your application.

It is also possible to start TextSoap and open an editable text document directly within TextSoap.
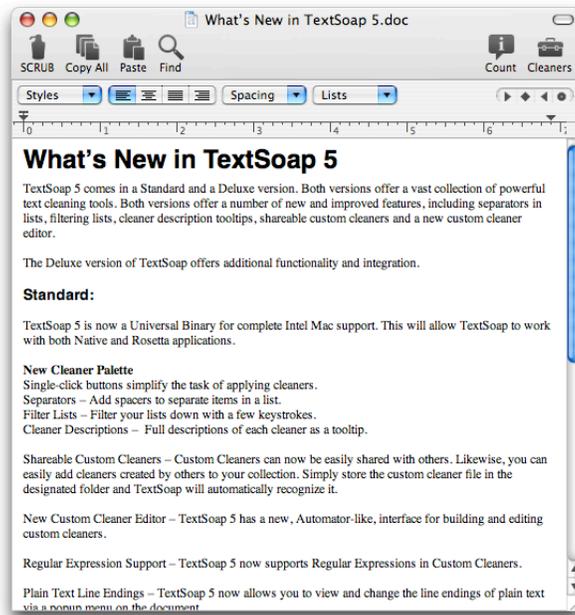
Option 2: Services Menu

Applications such as Apple Mail, BBEdit, Eudora, TextEdit, many Carbon and most Cocoa-based applications support OS X Services. TextSoap can be activated directly within a supported application through the Services submenu, of that application, for example "TextEdit".

1. Select text in application that supports Services (text must be editable).
2. Choose TextSoap cleaner from Application -> Services -> TextSoap 5 submenu.
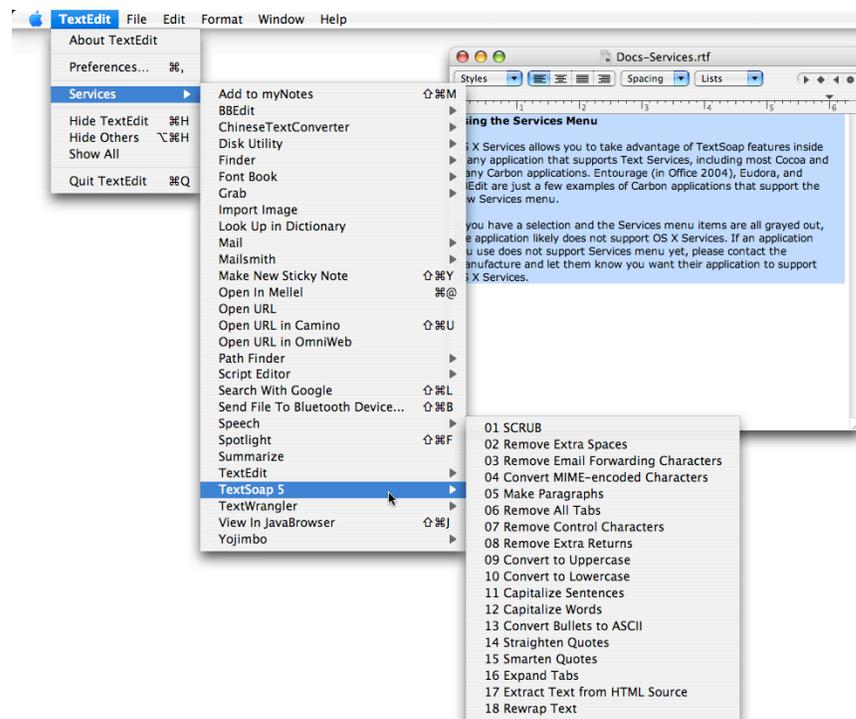
Option 3: Contextual Menu

Applications such as Entourage, Apple Mail, BBEdit, Eudora, TextEdit, and most Cocoa-based applications support contextual menus.
TextSoap can be started contextually within supported applications.

1. Select text in application that supports Contextual Menus
2. Control-Click to bring up contextual menu on selection
3. Choose TextSoap cleaner from TextSoap submenu
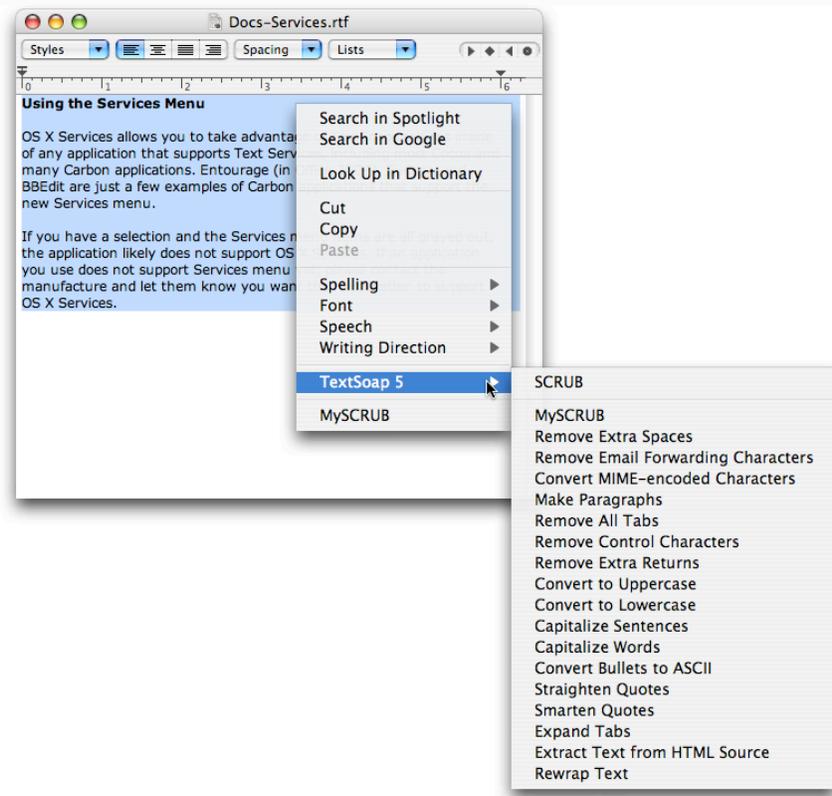
Option 4: Application Plug-in (Deluxe)

Requires a TextSoap Deluxe license.

Special Plug-ins for BBEdit, TextWrangler and Eudora are provided. The functionality varies based on the application.

1. Select text in application
2. BBEdit: Choose cleaner from palette
   Eudora: Choose menu from Edit -> Message Plug-ins submenu (or from toolbar if you configured it as such).

Whichever option you choose, the TextSoap helper tool applies the specified cleaner to your text.

See the application plug-in section for complete details.

The Cleaner Palette

The TextSoap 5 cleaner palette offers a number of key features designed to simplify your cleaning tasks by allowing for more task oriented grouping of cleaners, easy filtering of cleaners and providing the descriptions of cleaners without having to reference help or documentation.

Improved Single Click

No longer confined to a small area, the new cleaner palette makes each cleaner a one-click button.

Tell me what this Cleaner does

A frequently requested feature, TextSoap now displays the description of the cleaner as a tool tip.
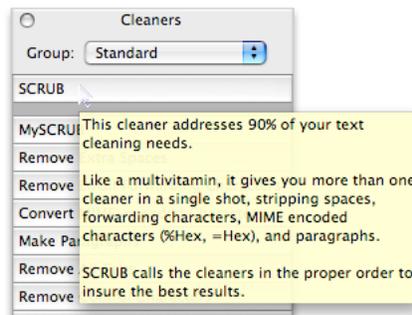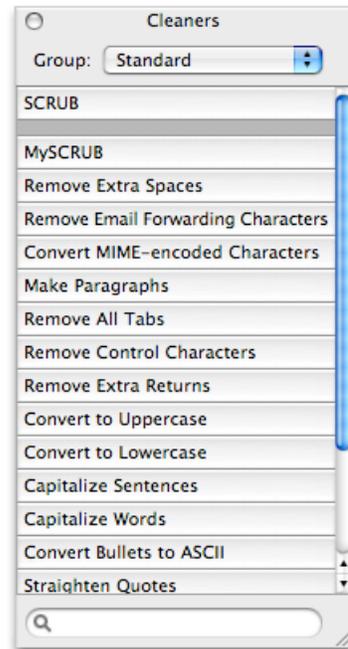
Finding with Filtering

TextSoap 5 allows you to filter the listing of cleaners within a group based on typing a few characters. An example:

1. Choose the Library Group of cleaners
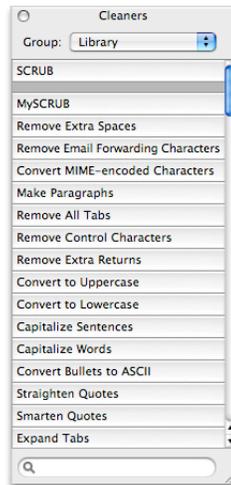
2. Click in the search box
3. Type in "Return"
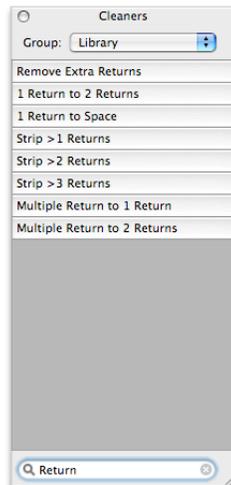
TextSoap will now show you all the cleaners in the group Library whose name includes "Return".

1. Click the X to clear the filter criteria

2. Now type in "quote"
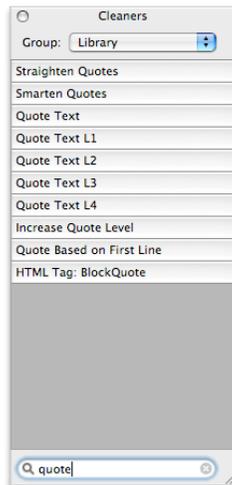
TextSoap will display the quote-related cleaners (having "quote" in the name). In this case, we have cleaners included that refer to quotes (as in quotation markers) as well as quoting (as in referring to some previous text).

Separated at Birth

TextSoap 5 now supports including separators in a cleaner list. This is a spacer item that visually separates a set of cleaners within a group. If you look at the default HTML group, you will see that separators are used to visually group common HTML tags together.
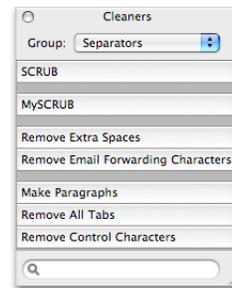
Compatibility Note: The size of the spacer varies. In OS X 10.3, the separator is the same height as a cleaner button. In OS X 10.4 or later, the separator is half the height of a cleaner button.

Grouping Cleaners (Deluxe)

Requires a TextSoap Deluxe license.

TextSoap 5 Standard offers a single customizable group of cleaners, called "MyList", to be used for your particular needs.

TextSoap 5 Deluxe expands on this by allowing you to create more than one group of cleaners. Groups can be used to better organize by the particular task you are trying to accomplish.
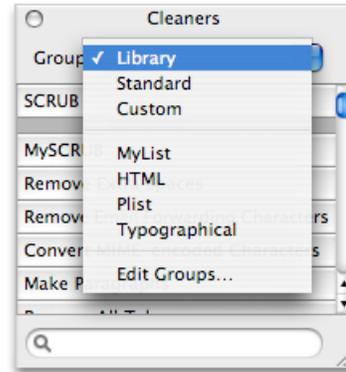
TextSoap 5 includes several pre-built task-oriented groups. You can customize and add your choice of cleaners to these as needed. It is only meant as a starting point for your own organization. To customize, simply choose "Edit Groups" from the Groups popup on the cleaner palette.

A group is a collection of cleaners that you define. An example might be cleaners you use often when processing email or perhaps when processing HTML text.

Let's create an Email Group.

1. After launching TextSoap, choose the cleaner palette group popup and select the "Edit Groups" item at the bottom of the menu.

2. Press the Plus button to create a new group.

3. Double click the "Untitled" item and rename it "Email".
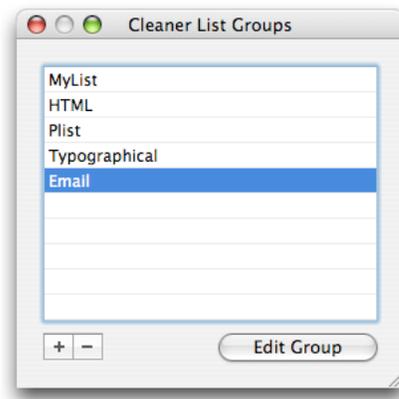
4. Select the Email group and choose "Edit Group". An editing sheet will appear allowing to build a list of cleaners you wish to be included in the group.
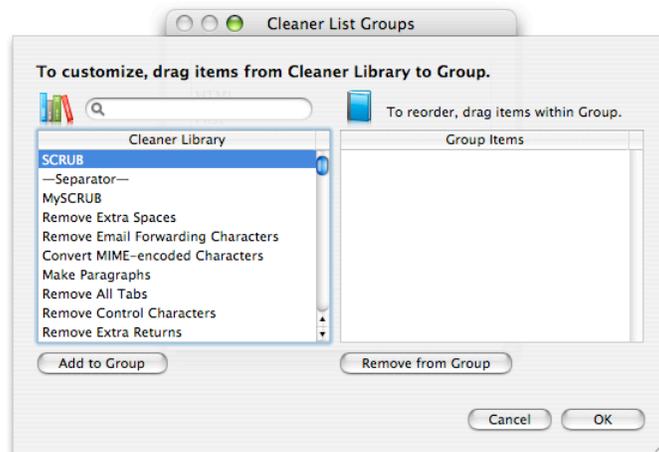
5. Select the "SCRUB" cleaner and choose Add to Group.

You can also drag-n-drop your desired cleaners from the library into the Group Items section.
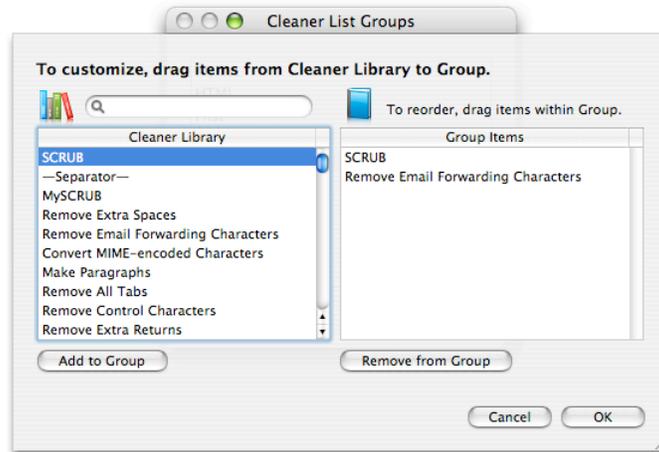
6. Now select the "Remove Email Forwarding Characters" and choose Add to Group.

7. Click on OK to accept the changes.

Now if you click on the Group popup menu and choose "Email", you should see the two cleaners you selected in your palette list.

If you would like additional cleaners in this group, you can edit it and add whatever other cleaners you might need.

You can create a virtually unlimited number of Groups to organize cleaners based on common tasks, document types or other criteria.
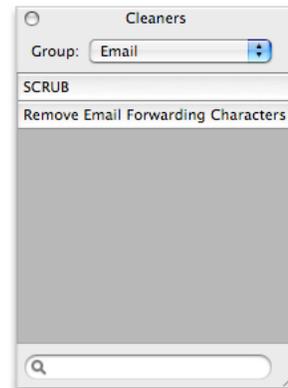
Integration Options

Using the Services Menu

OS X Services allows you to take advantage of TextSoap features inside of any application that supports Text Services, including most Cocoa and many Carbon applications. Entourage (Office 2004), Eudora, and BBEdit are just a few examples of Carbon applications that support the new Services menu.

If you have a selection of editable text and the Services menu items are all grayed out, the application likely does not support OS X Services. If an application you use does not support Services menu yet, please contact the manufacture and let them know you want their application to support OS X Services.

OS X Service menu items are derived from MyList set of cleaners, allowing you to customize the list of cleaners to pick from. To customize this list (or if there are no Service items visible), choose the "Edit Groups" from the Group popup and edit the "MyList" group to determine which cleaners you wish to have included in the Services menu.

TextSoap Deluxe allows you to create more than one custom group and allows you to specify which Group to use for the Services menu.

An Example: TextEdit

Let us assume that you have some text you want to clean that you've loaded up in TextEdit (a Cocoa application). Here's how to clean your text inside of TextEdit.

1. Select the text you want to clean.
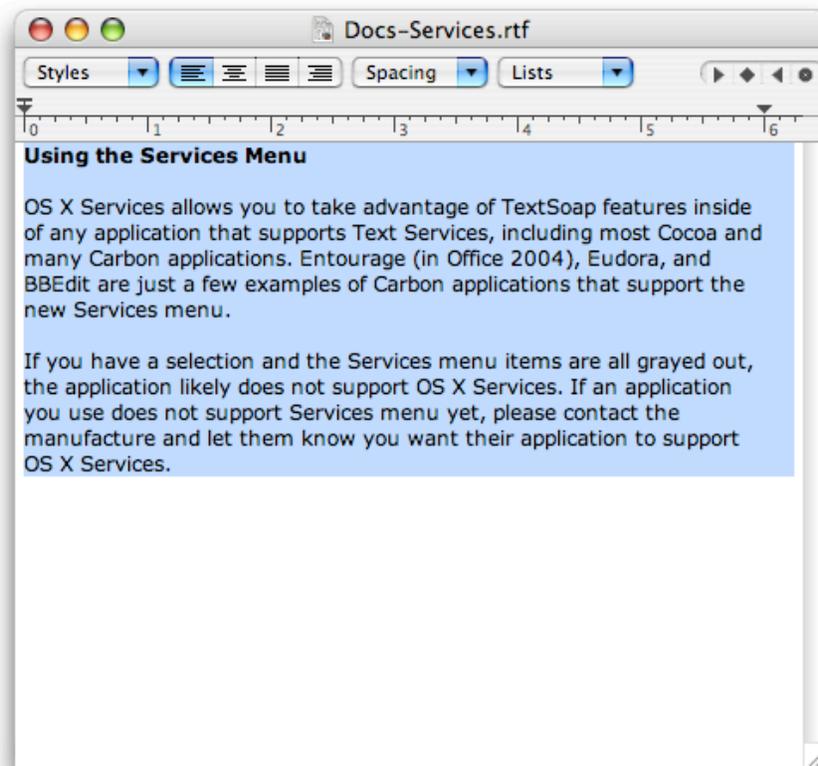
2. Go to the Application menu, Services sub-menu and then the TextSoap sub-menu and choose the cleaner option you want to use.
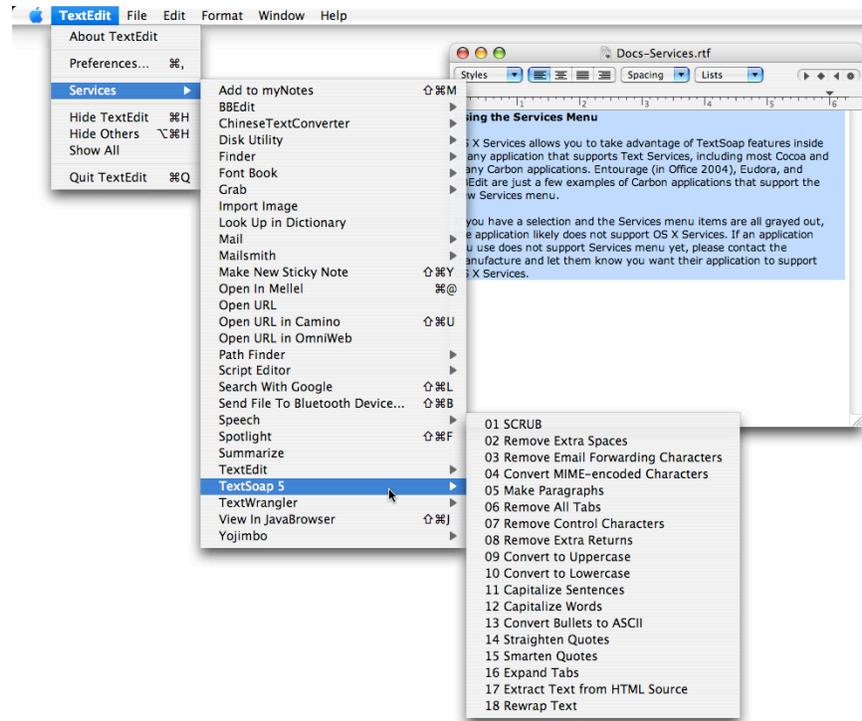
3. That's it!

If your not happy with the results, choose the standard undo and pick another cleaner.

Note: Because Services menus are sorted alphabetically, TextSoap places a two-digit number before the name to ensure the order for up to 99 items.

Using the Contextual Menu Plug-in (Deluxe)

Requires a TextSoap Deluxe license.

You can activate a Contextual menu inside an application by holding the control-key down while clicking the mouse (or if you have a two button mouse, clicking on the right button).

With Mac OS X, many applications now support Contextual Menu Modules, allowing this new functionality to be added into their application. Most Cocoa-based applications and many modern Carbon applications support contextual menus.

An Example: TextEdit.

Here's how to clean your text without leaving TextEdit (using the contextual menu).

Make sure you are in an editable document (you need to be able to change the text you want to clean).

1. Select the text you want to clean.
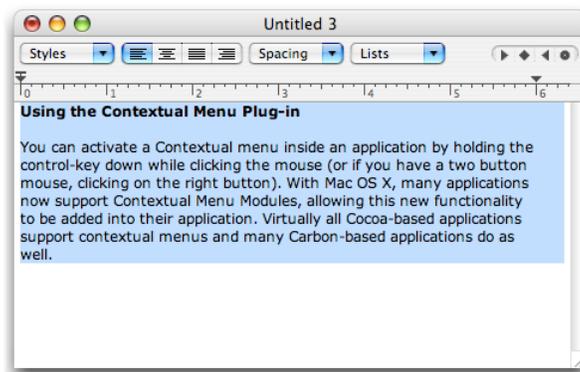
2. Control-Click (or Right-Click) to bring up the contextual menu.
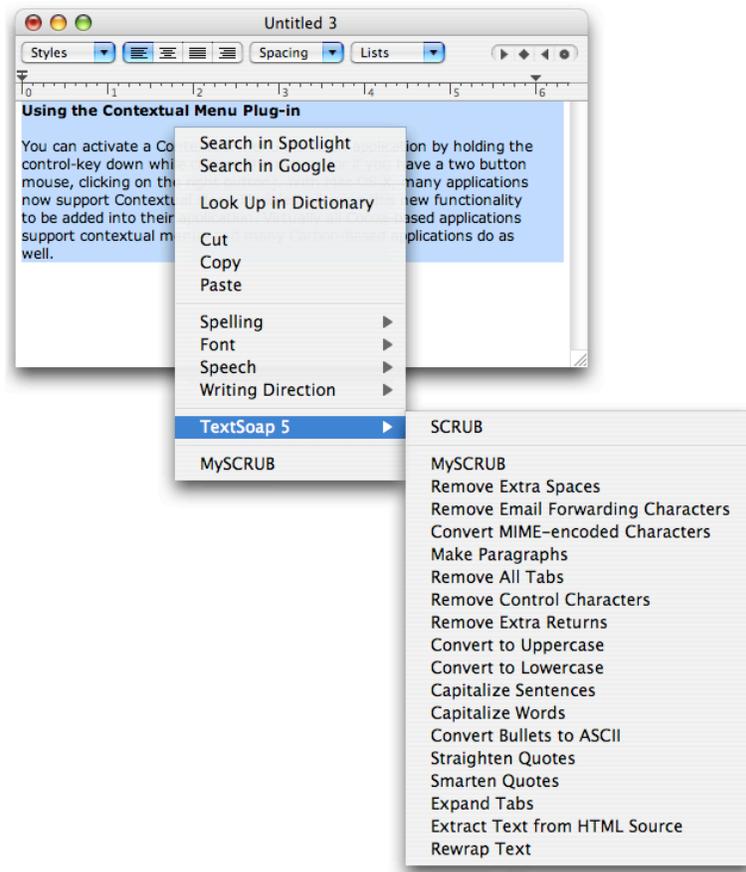
3. Choose the cleaner you want to use from the contextual menu.

That's it!

Note: Using the contextual menu plug-in will change the contents of your clipboard. When you clean text, the "cleaned" text is placed on the clipboard and the application is told to paste it. This works for most applications, but support is dependent on the host application.

Using the AppleScript Addition (Deluxe)

Requires a TextSoap Deluxe license.

An option available for those who want automated cleaning abilities via AppleScript. TextSoap has three standard commands available, one to clean text, one to clean the clipboard, and one to clean a file.

The syntax of the addition is as follows:

tsCleanFile *fileObject* with "*cleanerName*"

cleanerName–name of the cleaner (such as "SCRUB" or "Expand Tabs")
fileObject –file object variable – must be a text file

Note: Since clean text Files cannot be undone, files will not be cleaned in Trial Mode. TextSoap limits the number of characters that can be cleaned during trial mode.

tsCleanClipboard with "*cleanerName*"
cleanerName–name of the cleaner (such as "SCRUB" or "Expand Tabs")

tsCleanText *textObject* with "*cleanerName*"

cleanerName–name of the cleaner (such as "SCRUB" or "Expand Tabs")
textObject–text variable (such as a mail message) or any type of text you have.

Here is a sample that shows how to convert some text to uppercase using TextSoap.

```
on run
        set textData to 'sample text to convert"
        set textData to tsCleanText textData  with "Uppercase"
        display dialog textData
end run
```

Application Plug-ins (Deluxe)

Requires a TextSoap Deluxe license.

Installing

Go to the Preferences Window and choose "Plug-ins". The Plug-ins panel will indicate whether the plug-in is installed and if so, what version is running. Highlight the application plug-ins you wish to install and select "Install".

Using the BBEdit Plug-in (Deluxe)

The BBEdit plug-in allows you to take advantage of TextSoap features inside of BBEdit, TextWrangler, and Mailsmith. Here's how to clean your text inside of BBEdit.

If you don't have the palette visible, choose 'TextSoap 5 Tool" from the 'tools" menu.

1. Now select the text you want to clean (if you do not select any text, the entire document will be used).
2. Click on the cleaner button you want to use.

That's it!

If your not happy with the results, choose the undo (command-Z) and pick another cleaner.

The BBEdit palette now supports use of Groups and spacers. An experimental option allows the use of all the new TextSoap palette features (Groups, Spacers, Tool tips and filtering).

BBEdit also supports Services and Contextual Menus under Mac OS X.

Using the Eudora plug-in (Deluxe)

This plug-in allows you to take advantage of TextSoap features inside of Eudora. Let us assume that you have some text you want to clean that you've loaded up in Eudora. Here's how to clean your text inside of Eudora.

1. Select the text you want to clean.

2. Select the cleaner option you want to use from the Edit -> Message Plug-ins sub-menu

Note: Inside of Eudora, TextSoap cleaners have a "ts – " prefix.

3. That's it!

If your not happy with the results, choose the standard Eudora undo and pick another cleaner.

The Eudora plug-in will use the cleaners from MyList, allowing you to customize the list of cleaners to pick from.

Eudora also supports OS X Services and contextual menus.

Custom Cleaners

Custom cleaners are mini-scripts that allow you to define one or more steps to transform your specific text.

TextSoap offers a completely new Custom Cleaner Definition Editor.

Cleaners are now stored as individual documents in the in folder location:

~/Library/Application Support/TextSoap/Custom Cleaners/

where "~" is your home folder.

The steps available include:
- Apply Cleaner
- Find and Replace Text
- Find and Replace Repeating Character
- Rewrap Text
- Quote Text
- Insert Text

TextSoap Deluxe also adds:
- Tag Text
- Sort Lines
- Add Prefix to Lines
- Remove Prefix from Lines
- Add Suffix to Lines
- Remove Suffix from Lines

Apply Cleaner
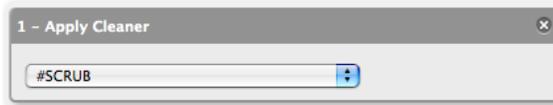
This step allows you to apply an existing cleaner to the selected text. The cleaner specified can be any built-in cleaner or custom cleaner.

Note: if the cleaner cannot be found, the command is simply ignored.

Find and Replace Text
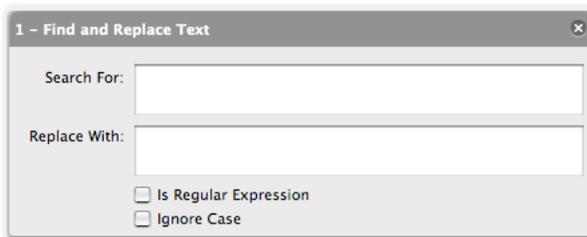
This step allows you to find a value and replace it with another value.

Is Regular Expression – check this if you wish to use regular expressions (See chapter on regular expression). If you don't know what regular expressions are, just leave this unchecked.

Ignore Case – check this option if you wish to find the text without regard to the case.

Find and Replace Repeating Character
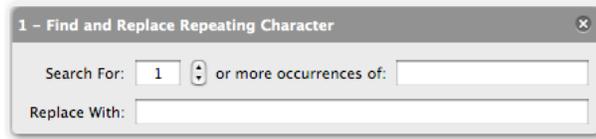
This step allows you to search for N or more occurrences of a character and replace it with a value. The search is only of a specific character (after the meta characters are converted).

For example, if you specify 2 or more "\t", the step will match tab-tab or tab-tab-tab.

Find and Replace (Special characters)

Find and Replace strings also allow you to specify some meta-characters, the table below describes each:

\\	=	backslash (\) character
\r	=	traditional carriage return character
\n	=	newline character
\f	=	formfeed character
\t	=	tab character
\x{HHHH}	=	Hexidecimal character with the hex value: {HHHH} (indicating a Unicode character)

Rewrap Text
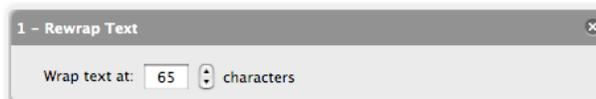
This step will rewrap the text at the specified number of characters. The text will be word wrapped to fit within the given number of characters.

Quote Text
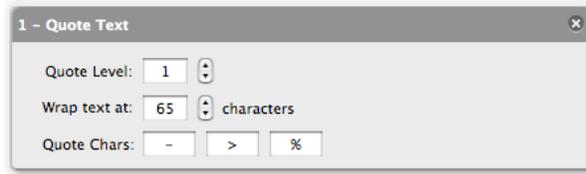
This step allows you to quote text. You can specify the

- Quote level
- Text wrapping
- Characters used to quote

For example, with a quote level of 4 and a quote marker of ">", the quoted string would be prefixed with a " >>>> ".

You can override the characters used to quote the text. You can specify a leading set of characters (usually spaces), the actual quote characters and the trailing characters (usually spaces).

Note: The actual wrap amount is this number minus the number of prefixes characters needed. In the example above, the actual text wrapping is 65 – 6 characters (leading space, four ">", trailing space).

Insert Text
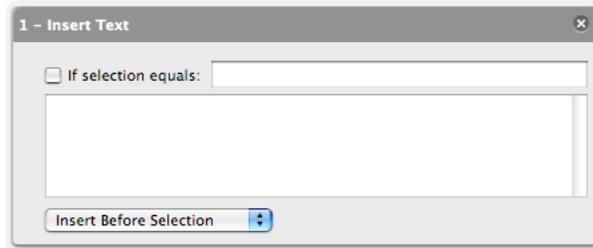
Clicking on the "Insert Text Options" button brings up a dialog that allows you to define the insertion of text in a cleaner. This is handy for inserting boilerplate text or, using the conditional test, expanding acronyms.

Checking the "If selection equals" item will enable a conditional text insertion. The selected text is tested against the string, if it matches, then the text insertion/replacement is done. If it does not match, the text insertion is skipped.

The third option allows you to specify how the text is inserted. It can be inserted before or after the selected text or you can use it to replace the selected text.

Requires a TextSoap Deluxe license.

The following steps are available for TextSoap Deluxe users.

Tag Text
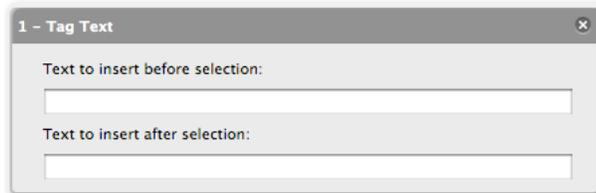
Specify the text to place before and after selected text.

An Example:

Specify the before tag as "<custom>" and the after tag as "</custom>".

Now select a word like "this" and apply the cleaner.

Now "this" becomes "<custom>this</custom>".

The HTML and PLIST tag cleaners are specific examples of this type of cleaner step in action.

Sort Lines

6 – Sort Lines ✕

Sort Lines in : [Ascending ▲▼] order

Note: Lines are determined by paragraph markers.
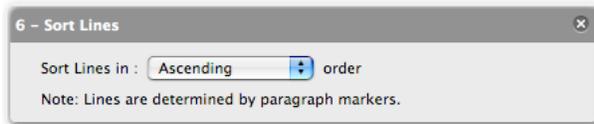
Specify which direction you wish to sort the lines.

For this cleaner step, lines are defined as ending with a paragraph marker.

Add Prefix to Lines

1 – Add Prefix to Lines ✕

Add this value to the beginning of each line:
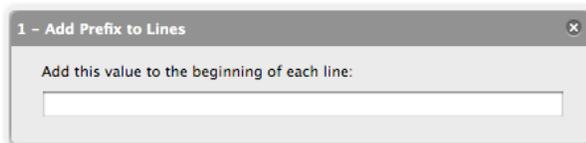
This allows you to add a prefix to the beginning of each line.

For this cleaner step, lines are defined as ending with a paragraph marker.

Add Suffix to Lines

2 – Add Suffix to Lines ✕

Append this value to the end of each line:
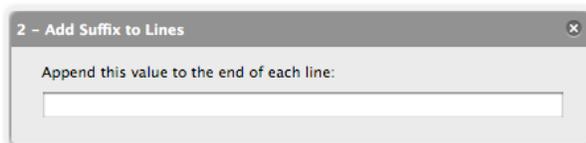
This allows you to add a suffix to the end of each line.

For this cleaner step, lines are defined as ending with a paragraph marker.

Remove Prefix from Lines
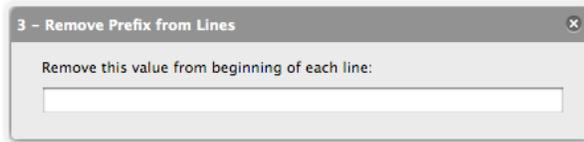
This allows you to remove a specific prefix (if it exists) from the beginning of each line.

For this cleaner step, lines are defined as ending with a paragraph marker.

Remove Suffix from Lines
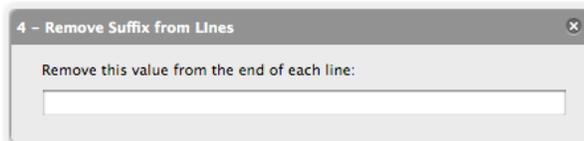
This allows you to remove a specific suffix (if it exists) from the end of each line.

For this cleaner step, lines are defined as ending with a paragraph marker.

Tutorial: My first custom cleaner

Now that you have the basics and as you become more familiar with TextSoap, you may want to create your own customized cleaner, set to perform to your personal specifications and/or preferences.

We are going to create two basic example custom cleaners using TextSoap.

First we will define our need. For this example, we want to be able to take raw HTML text, capitalize everything, remove HTML tags and rewrap content text to 40 character.

1. Create a new custom cleaner. Choose File -> New -> Custom Cleaner from the menu.
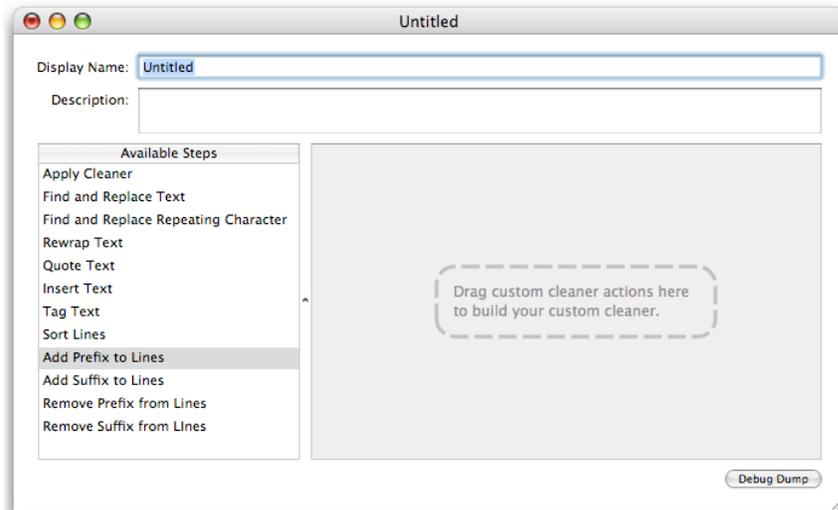
2. We give the cleaner a name and a brief description.
3. From the Available Steps, select and drag the "Apply Cleaner" step to the work area.
4. Choose "SCRUB" from the popup menu.
5. Repeat steps 3 and 4, adding the "Convert to Uppercase" and "Extract Text from HTML Source".
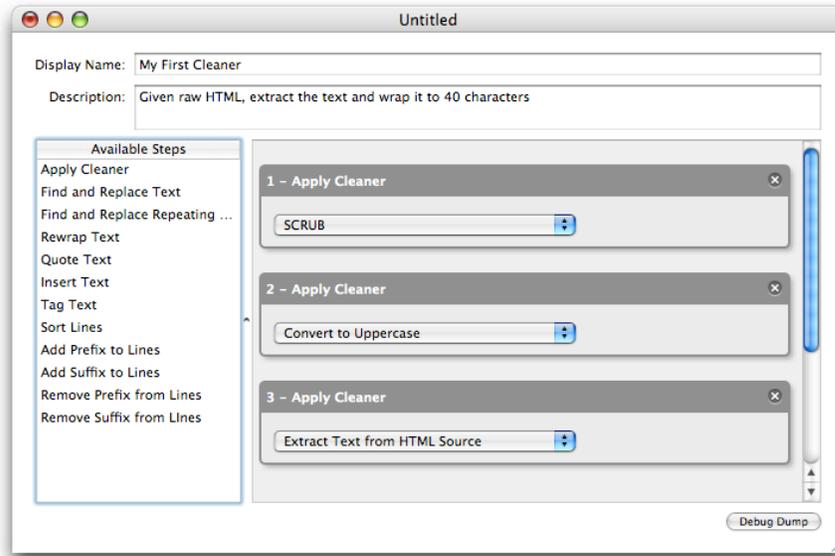
6. Now we drag a "Rewrap Text" step to the bottom of the work area and set the wrap amount to 40 characters.
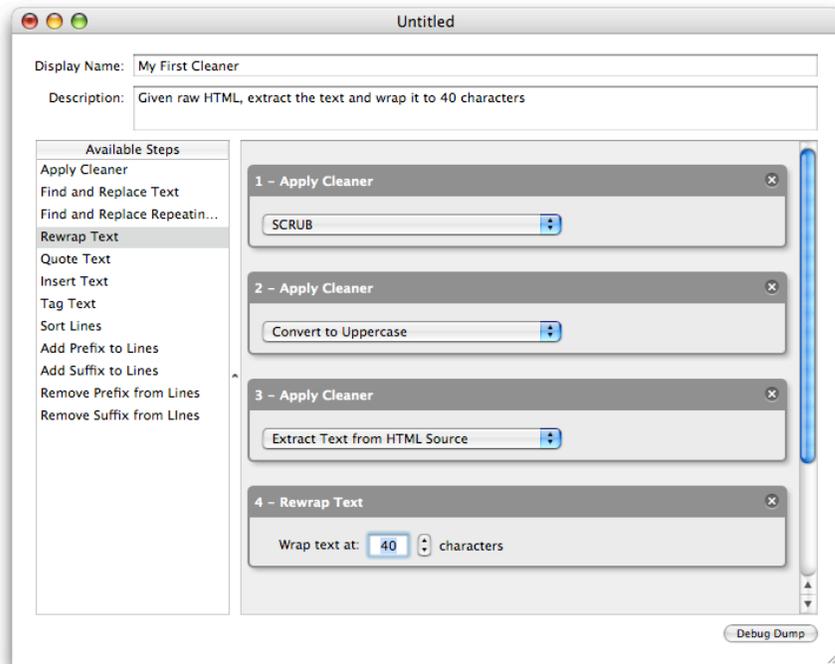
7. We save our custom cleaner. By default, TextSoap saves new custom cleaners to the appropriate folder.

Choose "Custom Cleaners" from the TextSoap application menu to see a list of all your custom cleaners.
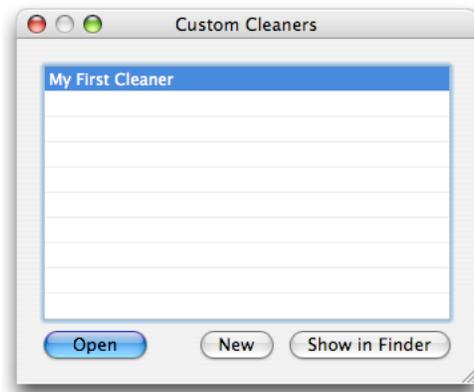
You can select your custom cleaner here to later edit it if you so choose or you can open it from within the Finder.

To view and apply the custom cleaner, you can choose either the "Custom" group or the "Library" group (which includes all cleaners).
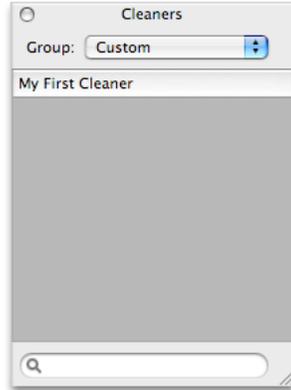
Custom Cleaners are basically mini-scripts or macros designed to save time on recurring tasks.

Creating custom cleaners is not for everyone. But one of the major new features of TextSoap 5 is the ability to share cleaners. Someone else can create a cleaner for you and you just need to drag it to the appropriate location. TextSoap will recognize you've added a new cleaner and include it in your list.

Regular Expressions

TextSoap now features regular expression support.

This section is not meant as a comprehensive discussion on using regular expressions. Regular Expressions are designed for very advanced users.

Not to worry though. TextSoap cleaners are now shareable, allowing you to use regular expression-based cleaners created by others.

Recommended Reading

For an in depth discussion on Regular Expressions, we recommend:

Mastering Regular Expressions, 2nd Edition
By Jeffery E.F. Friedl.
Publisher: O'Reily Media, Inc.
ISBN: 0596002890

TextSoap's Regular Expression support is based on OgreKit, using the Perl Compatible Regular Expression (PCRE) Syntax.

Note: There can be subtle differences between programs and the syntax that is supported.

What is a Regular Expression?

A regular expression is another way to describe a find or replace string.

When you choose to find something like "unmarked" you are asking to find a string that begins with a "u" followed by an "n" followed by an "m" followed by an "a" followed by an "r" followed by a "k" followed by an "e" followed by a "d". Wow! Luckily, the way we describe this is to just put the characters next to each other. Then, when all these conditions are met, the search string is considered found.

Regular expressions allow you to use special characters to further describe the search string. For example, we can use a regular expression to say string that begins with a "u" followed by a number of characters, but ends in a "d". The regular expression for this might be "u.*d". The two middle characters act as special characters used to represent "any number of characters".

Understanding Regular Expressions.

Most characters match themselves. The letter "u" will match "u" in text.

Escape Sequences can be used to match special characters that don't naturally match themselves. For example, dot (.) has special meaning to represent any character, but if you want to match an actual dot, you would specify "\." (backslash followed by a dot).

Reference

See the TextSoap Regular Expression Reference document for a description of the regular expression options.

Preferences

General
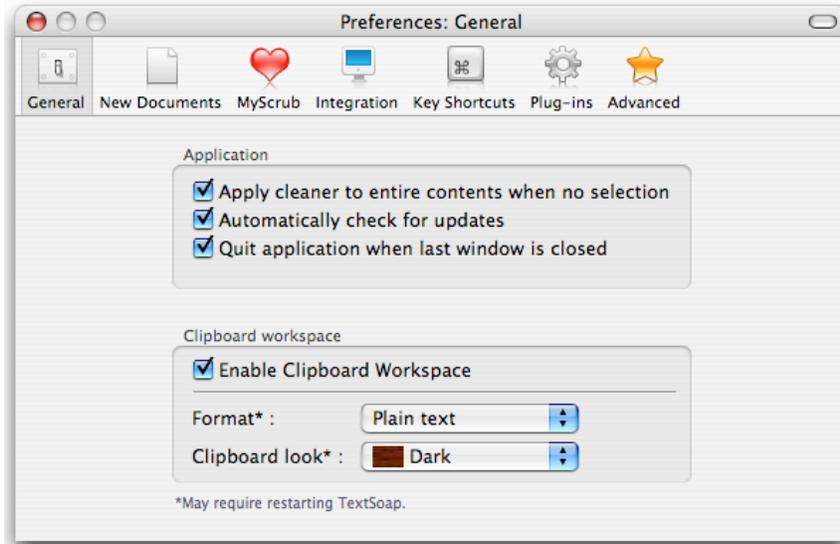
Apply Cleaner to entire Document if no selection – indicates whether to apply the cleaner to the entire document or just the selection. This option is off by default.

Automatically check for updates – indicates whether TextSoap should regularly check for new updates. This option is on by default.

Quit application when last window is closed – will automatically quit the application when you close the last document. This option is on by default.

Enable Clipboard Workspace – indicates whether to use the clipboard workspace and auto copy/paste or simply create a new untitled document.

Format: You can specify whether you want plain text or Rich text (with fonts, paragraph formatting, etc) as your clipboard document. This option is set to "Rich text" by default.

Clipboard look: Allows you to specify the look of the clipboard workspace. Several styles are available (changes between Original Style and New Style require re-launching TextSoap).

New Documents
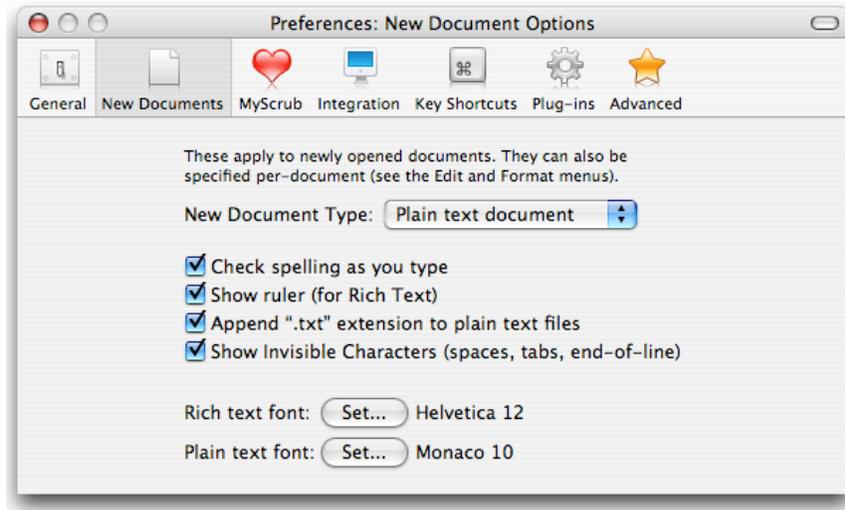
New Document Type – allows you to specify if you wish to create plain or rich text documents when you select New (or press command-N). This option is set to "Rich text document" by default.

Note: You can use Command-Option-N to always create a new Plain text document or use Command-Control-N to always create a new Rich text document.

Check spelling as you type – allows you to activate the automatic spell checker for new documents. The option is on by default.

Show ruler (for Rich Text) – allows you to indicate whether to show the ruler by default when creating a Rich Text document. The default for this option is on for rich text documents (the ruler does not display for plain text documents).

Append ".txt" extension to plain text files – standard option that mimics TextEdit's behavior. The option is on by default.

Show Invisible Characters (spaces, tabs, end-of-line) – shows spaces, tabs, returns and other invisible characters. You can toggle this per document using the toolbar button. The option is on by default.

Default Fonts – allows you to specify what font to use depending on the type of document created.

MyScrub
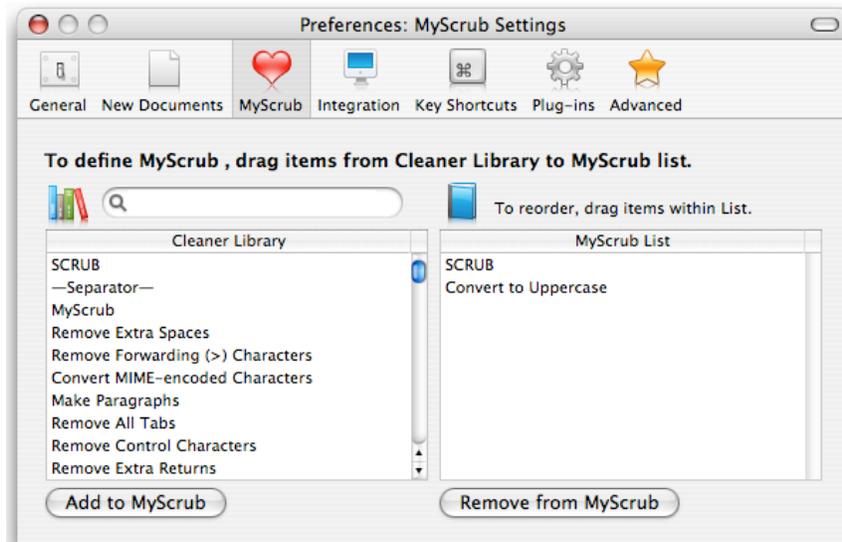
Define the cleaners you wish to use when you click on the MyScrub cleaner. This preference acts like a a very specialized "Custom Cleaner", with multiple Apply Cleaner steps (one for each cleaner listed).

System Integration
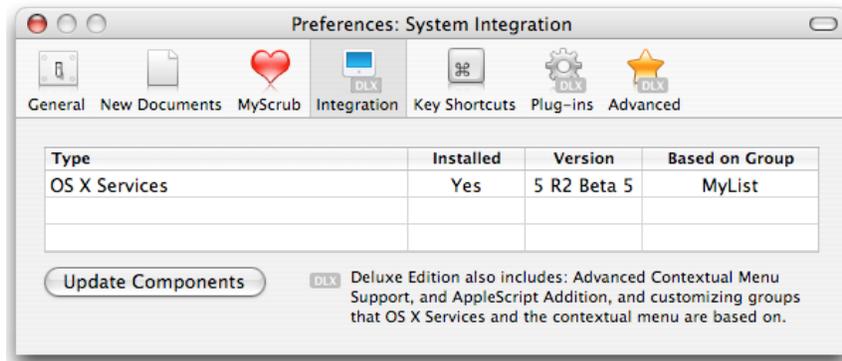
This preference shows you what system integration components are installed and allows you to force an update of the components. TextSoap Standard supports OS X Services.
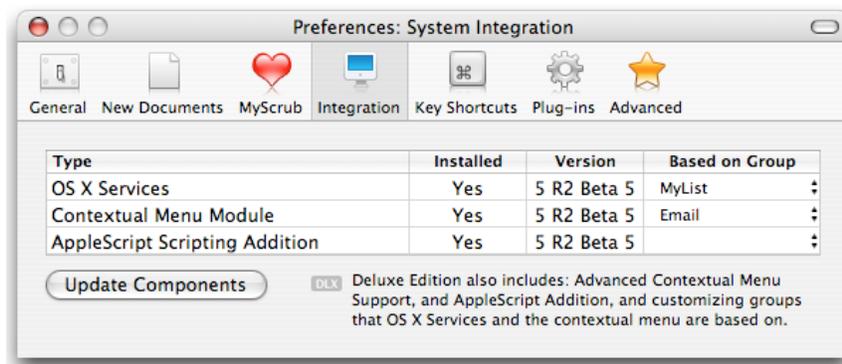
If you are using the *Deluxe* version, you have the Contextual Menu Module and AppleScript commands that are also installed on your system.

The Deluxe version also allows you to specify which Group to use for each of the plug-ins.

Key Shortcuts
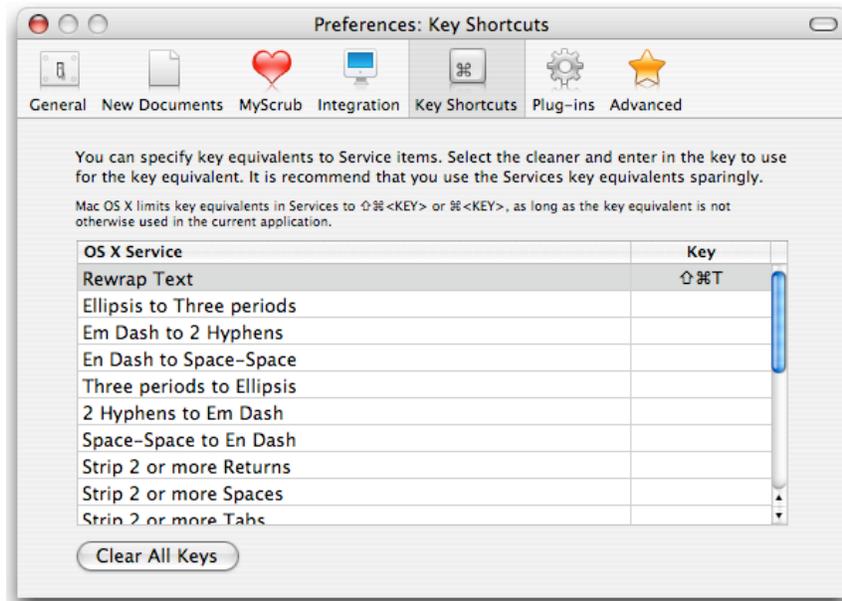
This preference option allows you to specify keyboard shortcuts for OS X Service items. OS X Services does set a few limitations on using key equivalents. It needs to be either ⇧⌘KEY or ⌘KEY where KEY is an ASCII character and it must not otherwise be used in the current application (we didn't make the rules).

To remove a key equivalent, select the key field and press SPACE.

TextSoap sets the key for the specific cleaner, so even if you change groups – if that cleaner is contained in the group, it will retain its key equivalent. TextSoap will only set the key equivalent for an item that is to be displayed in OS X Service menu. Key equivalents set for cleaners not in the group used by OS X Services will be ignored.

Because of these limitations, we recommend that you use key equivalents sparingly. *Tip: You can change the key equivalents "on the fly". If you have use for a specific cleaner, setup a key equivalent, in most cases the changes are immediate. When you're done, just remove it if necessary.*

Clear All Key Shortcuts – This option allows you to remove key shortcuts associated with all the cleaners.

Plug-ins (Deluxe)

Requires a TextSoap Deluxe license.
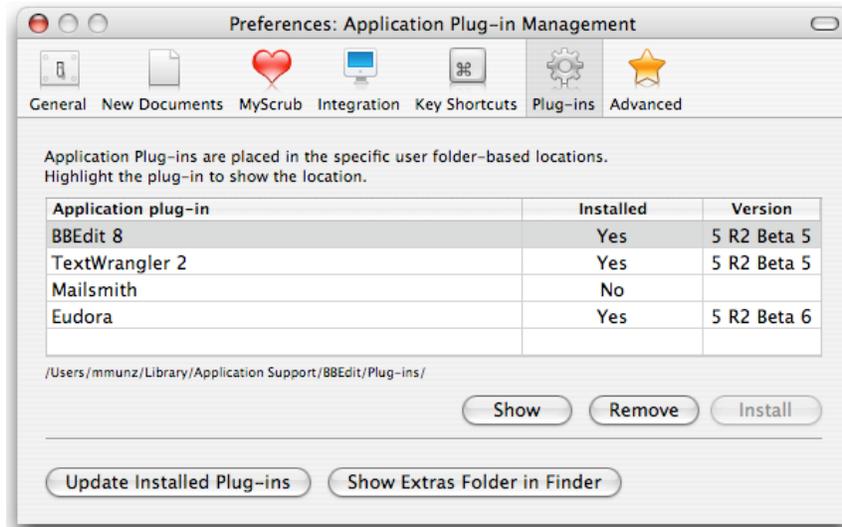
TextSoap 5 offers some new plug-in management tools to ease the installation and updating of these plug-ins. Each tool has a check box and shows you the location of the actual plug-in. System tools are installed by default and cannot be unchecked.

You can optionally install plug-ins for BBEdit, TextWrangler, Mailsmith and Eudora by checking the item next to the appropriate name.

BBEdit, TextWrangler and Mailsmith allow plug-ins in the Applications support folder. The plug-ins are copied and updated in this location. The applications will automatically pick up the updated versions.

Eudora does not support plug-ins in the Applications support folder. This requires one additional step by the user. After checking the items. An alias to the plug-in (now located in ~/Applications Support/Eudora/Plug-Ins/) is created and placed in the Eudora/Contents/PlugIns/ folder). Doing this allows the plug-in to be updated in the Applications Support folder and still recognized in the Eudora application

Advanced (Deluxe)
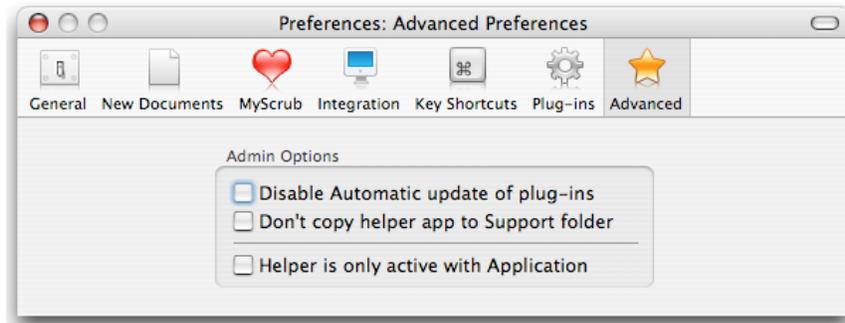
Administrators of machines with multiple users may find the following options helpful in maintaining the machines.

Disable automatic update of plug-ins – This option will disable the feature of automatically update any installed plug-ins when a different version of TextSoap is run.

Administrators that wish to manually install the plug-ins on a per-machine (vs. a per-user) basis can check this time.

Don't copy helper app to support folder – To support drag-n-drop install, TextSoap copies a helper application into Applications Support folder.

Administrators will want to turn this option on to run the helper application directly from the Application. This prevents TextSoap from making copies of the helper application on a per-user basis.

The one side effect of this is that you need to drag the old application to the trash and then copy in the new one. The Finder will not replace the application otherwise.

Helper is only active with Application – This option only keeps the Helper application running when the main application is running.

*Note: The Helper application is needed to support OS X Services, Contextual Menus, the application plug-ins or AppleScript Addition. These services will be **unavailable** if the helper application is not running.*

This option is off by default for registered users.
This option is on (and disabled) while the application is in trial mode.

48